

## The Clinical Utility of the Rorschach: Unfulfilled Promises and an Uncertain Future

John Hunsley  
University of Ottawa

J. Michael Bailey  
Northwestern University

The empirical evidence on the Rorschach is reviewed using three definitions of clinical utility: (a) the nature of professional attitudes and extent of clinical usage, (b) the extent of evidence for reliability, validity, diagnostic efficiency, and incremental validity, and (c) the extent of evidence that Rorschach data improve clinical decision-making and/or treatment outcome. Surveys demonstrate that the Rorschach is extensively used; however, these data are insufficient to demonstrate clinical utility as they do not address the rational, scientific, and ethical requirements of professional standards for psychological measures. After reviewing conceptual issues in Rorschach research (especially those in the Comprehensive System) the authors conclude that there is little scientific evidence to support the clinical utility of the Rorschach. Given the absence of data evaluating how the Rorschach is used in routine practice and whether its use is consistent with the manner in which it is used in research, there is currently no scientific basis for justifying the use of Rorschach scales in psychological assessments.

The Rorschach has the dubious distinction of being, simultaneously, the most cherished and the most reviled of all psychological assessment tools. Countless articles and chapters reviewing the Rorschach over the past 50 years have told the same story: The Rorschach is held in great esteem by many psychologists for its ability to access intrapsychic material, whereas others point to the Rorschach as a prime example of unscientific psychological assessment. In recent years, this depiction of the Rorschach has begun to shift somewhat, as authors have alluded to the possibility that the extensive work of Exner (e.g., Exner, 1993) has begun to address the psychometric and scientific concerns of even the most ardent Rorschach critics. Groth-Marnat (1997), for example, stated that Exner's reliance on empirical validation of Rorschach summary scores and his development of a large normative database have increased the Rorschach's acceptance and status as a psychological assessment instrument.

This longstanding debate about the scientific and professional status of the Rorschach has clearly affected claims made about the nature of the Rorschach and the quality of Rorschach research. Few Rorschach advocates, for example, now liken the test to an x-ray of the psyche. Although the Rorschach inkblot test is, after the MMPI/MMPI-2/MMPI-A, the most commonly researched psychological test (Butcher & Rouse, 1996), most of the early research suffered from severe design and analysis problems. Exner (1986), for example, estimated that two thirds of the research published prior to 1970 was so flawed that it could not be considered valid. The quality of Rorschach research has improved, as

more attention has been devoted to issues such as experimental design, interrater reliability, and statistical power. However, few would argue that the debate over the merits of the Rorschach has been resolved to the satisfaction of all. Indeed, little appears to have changed with respect to both the polarized views of the Rorschach held by various psychologists and the force with which both critics and advocates state their positions.

The Rorschach is a complex and time-consuming measure to use: The modal time for administration, scoring, and interpretation is 3 hr (Ball, Archer, & Imhof, 1994). Given the typical fees charged by psychologists, this places the Rorschach among the most expensive psychological tests to use. Nevertheless, on the basis of recent surveys, the Rorschach continues to be extensively used (e.g., Watkins, Campbell, Nieberding, & Hallmark, 1995) and is taught in most American Psychological Association-approved clinical training programs (Piotrowski & Zalewski, 1993). The results of these and other surveys also suggest that the Rorschach is frequently used as part of an assessment battery, most typically combining clinical interviews and self-report personality measures (especially the various forms of the MMPI; Ganellen, 1996c). It also continues to be frequently used in forensic work and in child custody evaluations (Ackerman & Ackerman, 1997; Lees-Haley, 1992).

Although Exner's Comprehensive System for administering, scoring, and interpreting the Rorschach has gained dominance in the research literature (Shontz & Green, 1992) and in graduate training on the Rorschach (Hilsenroth & Handler, 1995; Ritzler & Alter, 1986), the extent of its use in clinical settings is unclear. Historically, psychologists have tended to "augment" and "personalize" the scoring and interpretation of the Rorschach based on different scoring systems and their own clinical experience (e.g., Exner & Exner, 1972). It is likely that this trend continues, as many graduate courses involve training in multiple scoring systems and 25%–37% of Rorschach courses do not include instruction in the Comprehensive System (Hilsenroth & Handler, 1995).

---

John Hunsley, School of Psychology, University of Ottawa, Ottawa, Ontario, Canada; J. Michael Bailey, Department of Psychology, Northwestern University.

Correspondence concerning this article should be addressed to John Hunsley, School of Psychology, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5. Electronic mail may be sent to hunch@uottawa.ca.

Moreover, the same authors who have described Exner's efforts to systematize the Rorschach in laudatory terms also have tended implicitly or explicitly to encourage the use of the Comprehensive System in conjunction with qualitative analysis and/or scoring systems that have strong clinical appeal despite having very limited empirical support (e.g., Aronow, Reznikoff, & Moreland, 1995; Fischer, 1994; Hilsenroth & Handler, 1995; Shontz & Green, 1992; Weiner, 1994). In sum, there is no documented evidence indicating how Exner's attempt to provide a scientifically sound approach to the Rorschach has affected the manner in which the Rorschach is typically used by psychologists.

With these contextual issues in mind, the purpose of our review of the Rorschach literature is to examine the extent to which the Rorschach has been demonstrated to have clinical utility. Although there is frequent discussion of the concept of clinical utility in the assessment literature, there is little consensus on an exact definition. Therefore, for present purposes, we define clinical utility in the broadest manner possible by offering three separate (and increasingly stringent) definitions. Our first definition of clinical utility is the simplest, and is directly tied to the everyday world of the practitioner: Is there evidence that psychologists find the Rorschach to be of value in routine clinical practice? The second definition of clinical utility is more directly related to the research base of the Rorschach: Is there replicated empirical evidence that the Rorschach can contribute to clinical activities by (a) providing reliable and valid information about psychological functioning and/or personality structure, (b) aiding in diagnosis and differential diagnosis, and (c) demonstrating incremental validity, that is, providing clinical information beyond that obtained through other commonly used assessment strategies (e.g., life history data, base-rate information, other assessment measures)? Our third definition of clinical utility is the most stringent: As it is actually used in the field by psychologists, does the Rorschach typically improve clinical decision-making and/or treatment outcome?

#### Clinical Utility: Attitudes and Clinical Usage

In the 1960s, several surveys of academic clinical psychologists found that the perceived importance of the Rorschach was on the decline (Jackson & Wohl, 1966; McCully, 1965; Thelen, Varble, & Johnson, 1968). Nevertheless, survey evidence from the past 3 decades indicates that a large number of psychologists believe that the Rorschach is a valuable measure in clinical assessment. Although some variation is evident over time, many clinical psychologists in academic settings, in clinical training settings, and in other service delivery settings believe that clinical psychologists should receive training in the use of the Rorschach (Durand, Blanchard, & Mindell, 1988; Garfield & Kurtz, 1973; Pruitt, Smith, Thelen, & Lubin, 1985; Thelen et al., 1968; Wade & Baker, 1977; Watkins et al., 1995). Not surprisingly, therefore, surveys have also demonstrated that most students receive instruction and supervision in the use of the Rorschach during their academic and internship training (Durand et al., 1988; Hilsenroth & Handler, 1995; Ritzler & Alter, 1986; Ritzler & Del Gaudio, 1976).

Surveys of clinical psychologists have shown that the Rorschach is a frequently used test and that it is used by many clinical psychologists, at least occasionally, when conducting psychological assessments (Lubin, Larsen, & Matarazzo, 1984; Lubin, Wallis, & Paine, 1971; Piotrowski, Sherry, & Keller, 1985; Wade,

Baker, Morton, & Baker, 1978; Watkins et al., 1995). Behaviorally oriented psychologists are an exception to this pattern, as they are much less likely to use the Rorschach as part of an assessment (Elliott, Miltenberger, Kaster-Bundgaard, & Lumley, 1996; Piotrowski & Keller, 1984; Wade et al., 1978). Furthermore, compared to the 1980s, fewer clinical psychologists in the 1990s reported using projective tests, in general, as part of their clinical assessment activities (72% vs. 55%; Norcross, Karg, & Prochaska, 1997).

Many authors have described the ways in which assessment information derived from the Rorschach can contribute to the clinical enterprise. Weiner (1986), for instance, stated that Rorschach responses provide clues to the ways in which an individual deals with life experiences, including how the environment is perceived, how information is processed, and how the person copes with stress. Unfortunately, there is relatively little data on precisely what psychologists see as the utility of the Rorschach. In surveys of assessment practices, respondents have generally indicated that prior clinical experience with a test is a major determinant in the decision to use it. Furthermore, the primary motivations for using psychological tests have been as follows: (a) to answer specific assessment questions germane to diagnosis or personality structure and (b) to assist in treatment planning (Piotrowski et al., 1985; Wade & Baker, 1977; Watkins et al., 1995). Although extensive data have been collected regarding attitudes about and usage of the Rorschach, we know little about psychologists' perspectives regarding the unique contribution they believe the Rorschach makes to psychological assessment and treatment planning.

In sum, then, there can be little doubt that if a test's clinical utility is simply defined as the extent of its clinical usage, the Rorschach continues to have substantial clinical utility. In our opinion, however, this is an insufficient definition of clinical utility, because it ignores the rational, scientific, and ethical requirements inherent in professional standards for psychological measures (e.g., *Standards for Educational and Psychological Testing*, 1985). We thus turn to the more stringent, and in our view, more appropriate definitions of clinical utility.

#### Clinical Utility: The Scientific Perspective

In reviewing the Rorschach literature, we have been struck by the incongruity between the high regard many psychologists hold for the Rorschach and the many withering attacks on the empirical evidence typically adduced as providing scientific support for the Rorschach (e.g., Cronbach, 1949; Jensen, 1965). Relatedly, surveys of academic clinical psychologists in 1968 and 1983 indicated that, although the majority of respondents recommended that psychologists should know how to use the Rorschach, nearly two thirds indicated that they thought the empirical evidence did not support the value of the Rorschach and related techniques (Pruitt et al., 1985; Thelen et al., 1968).

Wade and Baker's (1977) survey of clinical psychologists is particularly illuminating. Examining expressed opinions about psychological tests, they found that many respondents were indifferent to negative research evidence about a test, choosing instead to rely on their personal experience. Similarly, clinical experience was more important in choosing tests and interpreting test results than were the test's psychometric properties. Many putative explanations have been offered concerning the limited relevance of

extant research for the clinical use of the Rorschach (e.g., Blatt, 1975; Howes, 1981; Levy & Orr, 1959; Schwartz & Lazar, 1979). Imperfections in the research notwithstanding, the apparent disregard of scientific evidence by many psychologists is troubling. Although it is uncertain whether these conclusions can be extrapolated to current assessment practices, there is no recent evidence to suggest that Wade and Baker's conclusions would not hold today. Moreover, given (a) the accumulating evidence of the negative effects of biases and heuristics on clinical judgment (e.g., Dawes, 1994; Turk & Salovey, 1988) and (b) the overwhelming evidence demonstrating the superiority of empirically derived decision-making rules over routine clinical judgment (Dawes, Faust, & Meehl, 1989), psychologists should be increasingly aware that assessment practices must be firmly grounded in empirical evidence.

In order to fairly evaluate the Rorschach from the scientific perspective inherent in our second definition of clinical utility, we focus our attention on the past 3 decades of research. Without doubt, the quality of research has improved in this period, largely due to researchers' responsiveness to earlier methodological criticisms and the increasing number of Rorschach researchers with strong methodological and statistical skills.

The increased use of Exner's Comprehensive System for administering, scoring, and interpreting the Rorschach has positively influenced the quality of Rorschach research. Although we will occasionally focus on a Rorschach scale not included in the Comprehensive System, we primarily focus our review on research using this system. Prior to Exner's systematizing efforts, the Rorschach was really many different tests, as there were five different scoring systems in common use (Exner, 1974). From these systems, Exner selected elements that had the most empirical support and combined them to form the basis of his Comprehensive System. Over the years, various editions documenting the Comprehensive System have included (a) detailed rules for administration, inquiry, scoring, and interpretation; (b) evidence of reliability and validity for many scales and summary scores; and (c) normative data for clinical and nonclinical samples. The current Comprehensive System is not compatible with the five systems that predated it, however, as there are important differences in (a) the administration of the Rorschach cards, (b) the inquiry performed by the psychologist in order to better understand the structural elements contributing to each response, and (c) the criteria used to score the responses. Attempts by psychologists to combine the current Comprehensive System with these other scoring systems are problematic, because variations in administration, inquiry, and scoring may result in significantly different patterns of scores (Blais, Norman, Quintar, & Herzog, 1995; Kinder, Brubaker, Ingram, & Reading, 1982; Ritzler & Nalesnik, 1990). Likewise, a substantial number of Rorschach studies cannot be adduced to support the Comprehensive System because of differences in test procedures (for a recent example in which the inquiry was conducted after each card, rather than after responses were obtained for all cards, see Khadivi, Wetzler, & Wilson, 1997).

### *Reliability Evidence*

In reviewing the psychometric qualities and scientific status of the Rorschach and the Comprehensive System, one must be cautious in referring to the test in a global manner. The Rorschach and

the Comprehensive System are actually compilations of dozens of scales, summary scores, and indices, and the reliability of each is heavily reliant on the scoring skills of the psychologist. Therefore, a thorough evaluation of the Rorschach's reliability requires separate consideration of each scale. Such an evaluation is beyond the scope of this article. With these caveats in mind though, it does appear that many of the scales central to the Comprehensive System can have adequate reliability. Evidence presented in Exner's multiple volumes on the Comprehensive System (e.g., 1993) and in reviews of the published research (Meyer, 1997a; Parker, Hanson, & Hunsley, 1988) generally supports the reliability (often test-retest reliability) of scales commonly used in research that are scored by trained raters. Moreover, since 1991, there has been an editorial requirement that Rorschach research submitted to the *Journal of Personality Assessment* have a minimum of 80% interrater agreement on the scoring (Weiner, 1991). Apparently, this has not affected the frequency with which Rorschach studies are published in the journal (Weiner, 1997). As we discuss later, there remains, however, the significant question of how reliably the Rorschach is scored in routine clinical practice (i.e., field reliability).

Although there is evidence that many of the main scales used in the Comprehensive System can be reliable, the adequacy of the extant research base is questionable. In particular, concerns have been raised about the method by which interrater reliability is typically calculated in Rorschach research. The following two main issues are relevant: the nature of the response base (i.e., the total Rorschach protocol or individual responses to each card) and the desirability of adjusting for chance agreement. Considerable confusion exists in the literature regarding the calculation of interrater reliability, but most researchers report some form of percentage agreement between raters. As recently discussed by several authors (McDowell & Acklin, 1996; Wood, Nezworski, & Stejskal, 1996a, 1996b, 1997), this is the least stringent form of interrater reliability and is likely to be inflated due to the lack of adjustment for chance agreement between raters. These commentators recommend the kappa statistic as the most appropriate measure of interrater reliability; in contrast, Meyer (1997a, 1997b) has recently argued that kappa is not appropriate for use with all Rorschach data, especially with indexes having low base rates. The manner (or manners) in which reliability should be calculated for Rorschach data must be resolved, and consensus must emerge on the requisite reliability criteria for publishing Rorschach research. Relatedly, there needs to be consensus on how to analyze data from indices with low base-rates, as the typical reliance on parametric statistics is not defensible (cf. Viglione, 1997).

### *Validity Evidence*

As with reliability, one must be cautious in making global statements about the validity of a test as complex and multifaceted as the Rorschach. It is common to see both Rorschach advocates and critics suggesting that the Rorschach, in toto, is either valid or invalid. Most recent publications summarizing the validity of Rorschach indices rely heavily on several meta-analytic reviews that purport to have found evidence of the validity of some Rorschach scales. It is therefore most appropriate that we begin our discussion of validity by focusing on these studies.

The meta-analyses reported by Parker (1983) and Atkinson (1986), based on a subset of the Rorschach literature, appear to offer support for the general validity of the Rorschach—however, because of a basic statistical problem, such a conclusion cannot be drawn. Both of these studies were conducted when there were few statistical guidelines for the conduct of meta-analytic reviews, and both used analysis of variance (ANOVA) statistics to analyze their effect size data. As outlined in one of the earliest guides to meta-analysis, there is no defensible argument for the use of ANOVA to analyze effect sizes or correlations (Hedges & Olkin, 1985). Meta-analytically derived data violate the homogeneity of variance assumptions underlying ANOVA statistics. As data from studies included in a meta-analysis are based on unequal sample sizes and have differing variances, any resultant *F* statistic will be incorrect. Accordingly, the Parker (1983) and Atkinson (1986) studies have historical merit for attempting to apply meta-analytic techniques to Rorschach data, but their results cannot be considered valid in light of subsequent developments in the use of meta-analysis.

Parker, Hanson, and Hunsley's (1988) meta-analysis avoided this statistical problem by using an appropriate analytic strategy (i.e., weighted regression analyses). Their findings suggest that some Rorschach indexes can possess moderate validity. The extent of the evidence stemming from this study has, however, been overstated by many Rorschach advocates, and conclusions that this meta-analysis proves the validity of Rorschach assessments (e.g., Shontz & Green, 1992; Weiner, 1997) are unwarranted. The Parker et al. (1988) meta-analysis focused on only nine scales, a small fraction of those in the Comprehensive System and an even smaller fraction of those studied in Rorschach research; furthermore, the database supporting the validity of these scales, five studies from a single journal, was extremely limited. Moreover, a recent reanalysis of Parker et al.'s data indicated that, in the sample of studies included in the meta-analysis, the typical validity of the Rorschach was significantly lower than that of the MMPI (Garb, Florio, & Grove, 1998).

Atkinson, Quarrington, Alp, and Cyr (1986) examined 120 Rorschach studies to determine the proportion of analyses yielding significant results compared to the total number of analyses conducted in each study. For conceptually guided studies, approximately half of the analyses were significant; among studies that lacked a conceptual rationale, only one fifth of analyses were significant. Although not based on effect sizes, these data, when combined with the results from Parker et al. (1988) do suggest that Rorschach data can yield scientifically valid information.

Are the global results from these two studies relevant to the issue of clinical utility? The charge given to contributors to this Special Section was to focus on evidence of direct relevance to the provision of psychological service (e.g., diagnosis, treatment planning, treatment outcome). To be consistent with this charge, we can only conclude that, given the very broad nature of these two studies and the range of studies included in their analyses, it is not possible to state that these results demonstrate validity of a form that is directly relevant to clinical practice.

With this in mind, we now turn to recent reviews of the evidence typically marshaled in support of the validity of the Comprehensive System and/or the Rorschach. Shontz and Green (1992), although often cited as providing evidence for the validity of the Rorschach, rely entirely on the Atkinson (1986), Atkinson et al.

(1986), Parker (1983), and Parker et al. (1988) studies to support their contention that the Rorschach is valid when used appropriately. In his 1996 article on the validity of the Rorschach, Weiner relied primarily on the meta-analytic research of Atkinson (1986) and Parker et al. Additionally, he indicated that Ornberg and Zalewski's (1994) review of 48 studies using the Rorschach with adolescents provided some evidence for the test's validity. A closer reading of this review article suggests that the extent of supporting evidence is rather limited. Of the 48 studies examined by Ornberg and Zalewski, only 10 did not suffer from what were termed "methodological concerns," and, of these 10 studies, only 6 were cited as providing support for the validity of the Rorschach. Finally, in a 1997 article on the current status of the Rorschach, Weiner used the Parker et al. and Shontz and Green (1992) articles to support his argument that the scientific merit of the Rorschach has been confirmed. When viewed together, it is clear that current assessments of the general validity of the Rorschach rely almost entirely on invalid or extremely limited meta-analytic evidence. As already stated, this meager evidence cannot be used to substantiate the clinical utility of the test.

In their concluding paragraph, Parker et al. (1988) called for more detailed meta-analyses examining specific subscales, populations, and dependent variables. With few exceptions, such as the recent meta-analysis by Meyer and Handler (1997), which we discuss later, this has not occurred for the Rorschach. This type of evidence is crucial for establishing both the specific validity of a Rorschach subscale and the overall validity of the Comprehensive System. A scale-by-scale review of validity is beyond the scope of this article and would probably require book-length treatment. Indeed, the several editions of the Comprehensive System might be cited as proof that such evidence exists for all of the Rorschach scales included in the Comprehensive System. Recently, however, Wood et al. (1996a) criticized much of the data presented by Exner in these editions. For example, they pointed out that, for some scales, (a) no supporting validity data are presented, or (b) research that does not support the validity of a scale is underemphasized, or (c) much of the empirical support comes from unpublished studies that have not undergone the usual scientific standard of peer review and have not been independently replicated (see also Kleiger, 1992). These criticisms are substantial enough to warrant a much more detailed presentation of the validity of the Comprehensive System scales, and the burden of proof clearly lies with the advocates of the System.

We now turn to some of the core methodological and statistical issues regarding the validity of the Comprehensive System. The most basic unresolved issue is the problem of response frequency. Respondents can provide multiple responses to each of the 10 cards, resulting in great variability in the number of responses to be scored. Although Exner has tried to reduce the effect of response frequency on the Comprehensive System by (a) adjusting scores for the number of responses, (b) calculating ratios and percentages, and (c) eliminating from analysis extremely brief protocols (fewer than 14 responses; Exner, 1988), problems still exist. For example, a number of Rorschach scales (e.g., Reflection and Food Responses) are based on a single response and thus are equivalent to a single item scale. Examinees who give more responses, overall, will have a higher likelihood of providing one of the scale-relevant responses. Meyer (1992a, 1992b, 1993) has reported that response frequency is significantly correlated with a number of scales and,

furthermore, on the basis of factor analytic research, factors defined by response frequency loadings may account for half of the explainable variance in Rorschach scales. As response frequency has been found to consistently relate to variables such as intelligence and educational level (Groth-Marnat, 1997), it is possible that the substantial effect of response frequency on numerous Rorschach scales is artifactual. The fact that half of the explainable variability in the dozens of Rorschach scales may be attributable to method artifact should certainly give pause to anyone claiming that specific Rorschach scales have been demonstrated to possess construct validity.

Despite a long history of such substantial problems, Exner (1992) continues to insist that response frequency is not a concern for the validity of the Comprehensive System. Indeed, several researchers have argued against strategies for systematically achieving a uniform number of responses across participants, as this might detract from the ideographic value of the Rorschach (e.g., Kinder, 1992; Lipgar, 1992). The reluctance of Exner and others to alter the Comprehensive System to require a uniform number of responses may be based in part on the understandable concern that psychologists would respond unfavorably to such a change in the Rorschach.

Demonstration of a measure's construct validity requires a thorough knowledge of the theoretical nature of the measure and an informed consideration of the validity criteria. For decades, the primary criteria against which the Rorschach has been judged have been self-report measures of psychological functioning and personality; for decades, researchers have usually found little relation between Rorschach scales and self-report measures purporting to embody the same constructs (e.g., Greenwald, 1990; Lipovsky, Finch, & Belter, 1989; Nezworski & Wood, 1995). Nowhere has the apparent lack of concordance been more evident than with the MMPI and MMPI-2. Part of the problem may be that, as with many areas of behavioral science research, much of the extant Rorschach research has limited statistical power; importantly though, research using the Comprehensive System is more powerful than other Rorschach research (Acklin, McDowell, & Orndoff, 1992). Another explanation for the problem involves the repeated finding that much Rorschach research has little grounding in theory or previous research: 24% of the reliability and validity analyses reviewed by Parker (1983) was exploratory in nature, 35% of the validity analyses reviewed by Atkinson et al. (1986) was exploratory, and 70% of the validity analyses reviewed by Parker et al. (1988) was exploratory. The finding that much MMPI research suffers from the same limitation may also contribute to the problem (Parker et al. found that 63% of MMPI validity analyses they reviewed was not grounded in theory or previous research and Atkinson [1986] found that 61% of MMPI validity analyses was exploratory).

Most research comparing the Rorschach and the MMPI/MMPI-2 has yielded little evidence of convergent validity (Archer & Krishnamurthy, 1993a, 1993b). Indeed, findings of minimal relationship between the two tests have been so common that at least 1 author suggested that the occasional significant findings may reflect Type II errors (Archer, 1996). Although a range of explanations for the lack of convergence has been offered by Rorschach proponents (Ganellen, 1996b; Meyer, 1996a; Nichols, 1996; Viglione, 1996; Weiner, 1993), these explanations are almost entirely post hoc. If the reasons why one should not expect

convergence were obvious, as some authors have argued, then it would be difficult to understand why dozens of studies have attempted to find hypothesized convergence between the tests. Some of the proffered explanations are, admittedly, quite intriguing, such as the possibility that the Rorschach may be better at assessing implicit, unconscious characteristics than at assessing characteristics within the respondent's awareness (Meyer, 1996a; cf. McClelland, Koestner, & Weinberger, 1989; Spangler, 1992). However, claims that the Rorschach's low-convergent validity reflects its virtues rather than its limitations must be supported by empirical evidence, not simply rhetoric.

Finally, in the Comprehensive System, the interpretation of the Rorschach data requires not only the calculation of numerous scales and ratios, but also the integration of the results from these various indexes. The interpretive meaning of any single scale is affected by the configuration of the other data (Exner, Viglione, & Gillespie, 1984). The combining and synthesizing of this information is far from a straightforward task, presenting the psychologist with precisely the type of task that is rife with problems in clinical judgment, as demonstrated by decades of research on biases and heuristics (Garb, 1998). The lack of evidence that different assessors interpret the entire database similarly is disconcerting and raises crucial questions about the validity of conclusions reached by psychologists. To be fair, the same can be (and has been) said of other complex psychological tests. However, the consistent finding that there is a relatively simple factor structure to the Rorschach test calls into question the relevance of the entire enterprise of scoring and integrating multiple indices. After accounting for the significant role played by response frequency, the main factors underlying Rorschach responses include variants of global psychological distress or health and affective control (Anderson & Dixon, 1993; Mason, Cohen, & Exner, 1985; Meyer, 1992b). There is, therefore, little reason to believe that the Rorschach effectively measures different facets of personality dynamics or personality structure—despite what generations of Rorschach proponents have claimed. There is also, therefore, little reason for psychologists to use strategies that differentially weight the results of the dozens of Rorschach scores when the Rorschach primarily appears to assess some form of broadly defined psychological adjustment.

### *Diagnostic Evidence*

In the Comprehensive System, several constellations of scores can be combined to form indexes of hypothesized relevance to clinical diagnosis, most notably the assessment of depression (the DEPI index) and schizophrenia (the SCZI index). On the sole basis of data presented by Exner (1991), Ganellen (1996a) concluded that these two indexes showed high diagnostic efficiency. Evidence published in peer-reviewed journals does not, however, support such a sanguine view of the DEPI and SCZI indexes. Although the SCZI index has been found to be significantly related to diagnoses of schizophrenia and other psychotic disorders in both adolescent and adult samples (Archer & Gordon, 1988; Meyer, 1993), all published research on the DEPI index indicates that it suffers from substantial sensitivity and specificity problems in relation to the diagnosis of depression (Archer & Krishnamurthy, 1997; Ball, Archer, Gordon, & French, 1991; Carlson, Kula, & St. Laurent, 1997; Meyer, 1993). Despite the ability of the SCZI index

to differentiate between psychotic and nonpsychotic disorders, it is less effective in doing so than the MMPI Scale 8 (Archer & Gordon, 1988). Overall, then, there is little evidence to recommend the clinical use of these Rorschach indexes for diagnostic purposes.

There is a further concern about the use of the Comprehensive System in the diagnostic process. Vincent and Harman (1991) applied the concept of clinical significance to some of the normative data for the Comprehensive System. Specifically they examined the data for nonpatient adults, and patients with depression, schizophrenia, and character problems. Using a standard of plus or minus two standard deviations from the nonpatient means, they examined whether the patient norms were different from those of the nonpatient population. These researchers reported that only 20% of 24 Comprehensive System scales attained the standard of clinical significance, and in most instances, clinical significant results were obtained only for the patients with schizophrenia. In essence, this suggests that, on the basis of Comprehensive System norms, few of the scales are able to distinguish between disordered and nondisordered adults, and even fewer are able to distinguish between nonpsychotically disordered and nondisordered adults.

#### *Evidence for Incremental Validity*

Rorschach proponents have suggested for decades that the best way to evaluate the clinical utility of the Rorschach is to examine its incremental validity (e.g., Widiger & Schilling, 1980), that is, its ability to aid in prediction above what can be obtained from other forms of data (e.g., base rates, demographics, life history, and other psychological tests). In light of accumulated evidence demonstrating little convergent validity between the Rorschach and relevant self-report measures, several Rorschach advocates have postulated that this provides an opportunity for the Rorschach to add important clinical data beyond that available from measures such as the MMPI (e.g., Acklin, 1993; Weiner, 1993). The limited evidence bearing on this question to date does not support, in general, the incremental validity of Rorschach scales. In a review of the incremental validity of personality assessment, Garb (1984) concluded that the addition of Rorschach data to demographic or self-report personality data never led to an increase in accuracy of personality assessments. It should be noted, though, that none of the studies reviewed by Garb specifically used the Comprehensive System for the Rorschach. In Archer and Gordon's (1988) study of the DEPI and SCZI indexes, adding data from these indexes to MMPI data did not increase diagnostic efficiency. Archer and Krishnamurthy (1997) recently obtained analogous results, in that Rorschach indexes did not improve upon the accuracy of MMPI-A indexes in diagnosing depression and conduct disorder. Thus, the frequent claim that the Rorschach adds meaningful assessment information to other data has not been supported in any published study. The Rorschach's incremental validity remains a tantalizing, unfulfilled promise.

#### *Summary*

In sum, there is little evidence to support the clinical utility of most Comprehensive System Rorschach scales when basic scientific criteria are applied to the published literature. One excellent example of a scale that does have scientific support (although not

part of the Comprehensive System) is the Rorschach Oral Dependency scale (Bornstein, 1996). The history of research efforts on this scale may serve as a useful guide for future attempts to validate Comprehensive System Rorschach scales, as it has a relatively large literature that includes varied research samples and varied measures of relevant validity criteria. Perhaps more importantly, research on this scale has also tended to be more focused than is typical in the Rorschach literature, with research designs constructed appropriately to test research hypotheses.

The reliability and validity of the Comprehensive System have been greatly overstated. The overreliance on unpublished research, the limited nature of current meta-analytic results, the paucity of replicated evidence, the questionable standards used for evaluating reliability, the problem of variation in response frequency, and the nature of the factor structure of Comprehensive System scales weakens the claims of many that the Comprehensive System has finally legitimized the Rorschach (cf. Wood et al., 1996a). Meyer (1996b) recently commented that, because of a history of criticism, Rorschach scales may need to meet a higher standard of scientific rigor in order to gain acceptance. In our opinion, critics do not require this higher standard; rather, they require simply that *each* Rorschach scale demonstrates the same level of reliability and validity as would be required of other psychological tests that may be deemed to have adequate psychometric properties. At present, the Comprehensive System, as a whole, does not meet the requirements set out in professional standards of practice such as the *Standards for Educational and Psychological Testing* (1985).

#### *Clinical Utility: Decision Making and Clinical Outcomes*

Our final definition of clinical utility is very stringent, requiring that in practice, Rorschach data contribute to better clinical decisions that have positive effects on people's lives. This definition is especially stringent for at least two reasons. First, it focuses on the actual use of Rorschach in clinical settings rather than in careful, controlled validity studies. Second, it requires not only that the Rorschach improve decision making, but that relevant decisions further the attainment of some commonly agreed upon desirable outcome. This definition is akin to Hayes, Nelson, and Jarrett's (1987) definition of treatment utility, as the focus is on whether the Rorschach makes a meaningful, desirable difference in treatment planning and resultant changes in psychological functioning and/or personality demonstrated by clients. Alternatively, in a nontherapeutic setting (e.g., a child custody evaluation), the use of a Rorschach must result in better decisions. A complete analysis by this definition would entail weighing the costs of administering the Rorschach against the benefits derived from its use.

Rorschach research on this strongest form of clinical utility has not approached adequacy, although to be fair, with the possible exception of ability tests in personnel selection (Hunter & Schmidt, 1996), neither has research about most other psychometric tests. Because no personality test has been examined with respect to this strongest form of utility, there is no empirical basis for comparison among tests. One possible reaction to this empirical lacuna is to ignore this most stringent definition and focus only on the forms of clinical utility for which evidence exists. We reject this position, because in our opinion, the neglect of strong clinical utility in the psychometrics literature must not be used to excuse the Rorschach or any other test. Rather, without relevant evidence,

there is no reason to assume that a test's use leads to good, rather than its opposite. Without evidence for this strong form of clinical validity, the profession or society in general has no good reason to encourage or accept an instrument's usage in its mental health and legal systems. Restricting our review to the weaker forms of clinical utility because only they have been studied would be akin to looking for lost keys under the street light merely because the light is brighter there.

Consider, for example, the question of whether Rorschach assessments should be reimbursable under managed care. Clearly, this decision should depend only on whether, broadly defined, the benefit to clients exceeds the cost, and not on whether the Rorschach can do something as well as, or better than, other instruments such as the MMPI. We do not believe that clinicians' opinions that the Rorschach is helpful should be sufficient to justify reimbursement for Rorschach assessments. There is no reason, in principle, why requisite research on the costs and benefits of Rorschach assessment (or assessment in general) cannot be done. It would be relatively straightforward, for example, to examine how much better outcomes are for clients assessed with the Rorschach compared with those who are not. It is precisely because assessment controversies have focused on the "trees" of relative reliability and validity rather than the "forest" of clinical utility that so little relevant evidence exists. If and when relevant research is performed, how is the Rorschach likely to fare? As discussed below, there are issues both general to assessment (e.g., the neglect of base rates) and specific to the Rorschach (e.g., high costs, the unstandardized way in which the Rorschach can be used) that make us doubt that the Rorschach is clinically useful by our stringent, but essential, criteria.

### *The Rorschach in Practice*

A careful study indicating that the Rorschach can have incremental validity in certain contexts by no means establishes that it typically does so in practice. There is a fundamental difference between the typical validity study and the clinical context. In the paradigmatic validity study, a researcher demonstrates that two carefully selected groups who differ on some psychological or behavioral attribute (e.g., thought disorder) also differ, on average, on a Rorschach score. Validity in the research context is typically reported as an effect size,  $r$  or  $d$ . If the effect size reliably exceeds zero, then the test has validity for the researcher. In contrast, the psychologist must use the information from the study in order to make judgments about particular clients (e.g., is the client thought disordered?). Validity for the psychologist must reflect the probability that a decision made on the basis of the Rorschach score is true. For the test to be valid, this probability must exceed the probability that an accurate decision would be made without the Rorschach. The following several factors could cause a Rorschach score to be incrementally valid in the research study but not in clinical practice: (a) reliance in the clinical setting on clinical rather than actuarial methods of combining information for prediction, (b) neglect of base rates in the clinical setting, (c) using the test on clients who are less prototypic than those used in the research study, and (d) administering or scoring the test differently in the clinical context than in the study.

Regarding (a), it is clear that combining information from different scales based on intuition is inferior to combining such

information using statistical formulas (Dawes et al., 1989), and a scale with incremental validity obtained using optimal weighting (as in multiple regression) will have less validity when the psychologist decides how to weight the different scales. Regarding (b), Meehl and Rosen (1955) showed that as base rates depart from 50%, predictions made using cutting scores derived from typical validity studies (which usually contain similar numbers of participants in two groups and hence approximately a 50% base rate of the characteristic to be predicted) diminish in accuracy. In some realistic examples, use of an instrument with moderate validity in the research context can even hurt prediction in the applied setting. For example, the use of the SCZI index to diagnose thought disorder or schizophrenia in settings with base rates of those problems as low as 10%–20% is probably inadvisable. Regarding (c), Kendell (1989) has argued persuasively that clinical validity will be reduced whenever findings from carefully selected, homogeneous samples are applied to "messier" clinical populations. We note that issues (a)–(c) are not unique to the Rorschach but apply to all psychological assessment instruments with which we are familiar. They are, nonetheless, quite important, and the nearly universal failure of test designers and users to address them casts doubt on the clinical utility of even scales with solid laboratory research support.

There is reason to suspect that the fourth issue we raised about generalizability of Rorschach research to clinical settings, that is, the alteration of administration or scoring procedures, is especially problematic. We have already noted that much of the existing enthusiasm for the Rorschach's scientific standing derives from research using Exner's Comprehensive System. Obviously, such research can reflect positively on clinical Rorschach practice only if the Comprehensive System (or something very similar to it) is capably used in clinical settings. There are several reasons to question whether this routinely occurs.

First, the Comprehensive System is a complex administration and scoring system, and any evidence of scoring reliability in research contexts would not automatically generalize to clinical practice. To develop and maintain proficiency in the Comprehensive System requires enormous time and effort. At present, we have no evidence for the field reliability of the system. Meyer (1997a) correctly claimed that even poor field reliability would have little bearing on the scoring principles of the Comprehensive System; however, poor field reliability does have crucial implications for the field use of the system. All psychological tests can be inaccurately scored, even structured questionnaires, and these scoring errors can easily result in interpretive errors (e.g., Allard, Butler, Faust, & Shea, 1995). Furthermore, there is extensive evidence indicating that the numerous contextual factors (e.g., appearance of the examiner or physical nature of the testing environment) may affect the nature of examinees' Rorschach responses (Masling, 1960, 1992). Thus, without compelling evidence that a Rorschach protocol is properly administered and scored according to the rules of the Comprehensive System, the validity of the interpretation arising from the scoring cannot be assumed.

The second reason for questioning the generalizability of the research stems from the long-standing tendency for psychologists to score Rorschach protocols in an idiosyncratic manner. Indeed, some authors emphasize that the most important data generated by a Rorschach can only be obtained by psychologists who bring all



of their experiences and acumen to understanding the content of clients' responses (e.g., DeCato, 1993). Although all psychologists have heard stories of the amazing prowess of Rorschach "wizards," the opportunities for mistaken interpretations to occur with unstructured and unsystematic scoring are readily apparent. The evidence to date indicates that the extent of clinical experience with the Rorschach is unrelated to the validity of test interpretations (Garb, 1989). Moreover, attempts to augment the scoring of the protocol by adding scales from other systems are misplaced, for scoring systems are not necessarily interchangeable, and scales should not be added by a psychologist to the scoring of a Comprehensive System protocol without evidence of reliability and validity.

A third reason for questioning the generalizability of empirical findings on the Comprehensive System to clinical practice is that there remains substantial disagreement among Rorschach advocates regarding the advisability of using the system at all. Exner's efforts to systematize the scoring of the Rorschach, as well as his continuing insistence that it does not assess projection (Exner, 1989), have alienated many proponents of the Rorschach (e.g., Aronow et al., 1995; Kramer, 1991). What is often not discussed in chapters or books summarizing the Comprehensive System is the extent to which psychologists may resist adopting the system because they disagree with attempts to turn the Rorschach into a psychometric test.

Our fourth reason for generalizability concerns is based on the recent revival of the argument that the Rorschach is not a psychological test at all, but that it is, rather, a method of interviewing that generates data relevant to the practice of clinical assessment (e.g., Weiner, 1994, 1997; for an earlier statement of a similar position, see Aronow & Reznikoff, 1972). Such a position appears to allow a Rorschach assessor to claim that there is scientific evidence supporting the Rorschach method while simultaneously freeing the assessor to use the data in a manner unconstrained by issues of scoring, norms, reliability, or validity. Ackerman's (1995) recent guide to conducting child custody evaluations provides a clear example of this stance. After stating that a Rorschach protocol should be scored with the Comprehensive System to avoid violating ethical standards, he then suggested that it is not always necessary to score a Rorschach protocol and that an experienced clinician can assess anxiety, depression, and thought disorder on the Rorschach without going through the rigor of formal scoring (p. 116). We fear that this approach to the Rorschach all too frequently describes how Rorschach data are used in practice. To his credit, Exner (1997) has adamantly insisted that the Rorschach is a psychological test by any traditional definition of the term and that formal scoring procedures must be followed.

### *The Rorschach and Clinical Outcomes*

Even if the Rorschach assessed certain things especially well in practice (and we think it is conceivable, although not yet demonstrated, that it does), this would not be sufficient to justify its extensive use. It is also important that incremental validity due to the Rorschach facilitates desirable clinical outcomes. In a treatment context, this would mean that on average, clients given a Rorschach (e.g., during a diagnostic assessment) ultimately should reach higher levels of functioning compared with those who were not tested with the instrument. This would raise the possibility of

selectively screening clients for treatments and/or developing alternative treatment strategies for those clients who appeared likely to have a poor prognosis. Other contexts may have different desired outcomes. For example, child custody examinations are primarily motivated by the goal of placing the child with the parent who will provide the best environment. In this context, the Rorschach's utility would be supported by showing that better placements are made with the instrument than without it.

The clinical utility of the Rorschach in the treatment context presumably derives from the effect of information on the treatment provided. In many settings, clients are interviewed, assessed, and diagnosed prior to treatment. Although many psychologists apparently place great value on accurate personality assessment, it is not obvious to us that this goal itself has utility. Research has generally failed to demonstrate that matching client psychological characteristics to treatment characteristics improves treatment outcome (e.g., Smith & Sechrest, 1991). Obviously, when a clinical goal is of questionable utility, the Rorschach, per se, cannot be faulted. Any assessment instrument used to achieve such a goal is equally suspect, because regardless of whether it uniquely contributes to the goal, it may be contributing no net benefit. As cogently argued by Hayes et al. (1987), assessment must be judged by its effects on treatment outcome.

To our knowledge, no study has examined whether Rorschach-informed pretreatment assessment leads to more effective treatment. There are, however, a number of studies that have examined the ability of various Rorschach scales to predict treatment-relevant parameters such as attrition and outcome in therapy (typically psychodynamic therapy). In general, these studies are neither systematic nor cumulative, as researchers have paid little attention to the importance of replication. Moreover, such studies are not immune from the methodological difficulties endemic to Rorschach research. For example, Hilsenroth, Handler, Toman, and Padawer (1995) recently presented data on the ability of several Comprehensive System scales to predict premature termination from long-term psychodynamic therapy. Of the nine scales chosen on the basis of previous empirical findings, only the three variables measuring interpersonal qualities were able to significantly distinguish between clients who did and did not terminate prematurely (defined as termination before the eighth therapy session). It is noteworthy that, of the nine scales, the distributions of scores on only these three interpersonal variables showed pronounced skewness, with four of six group means below 1.0 and standard deviations that were as much as twice the magnitude of the group mean. As is commonly acknowledged in psychological research, and as been repeatedly stated in the Rorschach literature (e.g., Kinder, 1992; Viglione, 1997), non-normal distributions should either be transformed prior to analysis or analyzed with nonparametric statistics. Unfortunately, the skewed data presented by Hilsenroth et al. were not transformed, and an ANOVA was used to compare groups. It is possible that their results are accurate, as an ANOVA is relatively robust in handling skewed data. However, predictions regarding the possibility of premature termination in other samples should not be made on the basis of these results, as cutoff scores derived from such skewed data are unlikely to be invariant or broadly applicable.

The Rorschach Prognostic Rating Scale (RPRS; Klopfer, Kirkner, Wisham, & Baker, 1951) is the sole Rorschach scale relevant to the prediction of treatment outcome that has been



repeatedly studied by independent groups of researchers. In a recent meta-analysis of this literature, Meyer and Handler (1997) reported that the correlation between the RPRS scale and outcome was .44. Although there are many merits to the analyses conducted by Meyer and Handler, problems with their sample of studies weaken their claims in support of the validity of the RPRS. For example, of the 22 statistics resulting from the researchers' literature search on the ability of the RPRS to predict outcome, 3 were not based on published studies of psychotherapy and 4 were based on samples that were not independent of those reported in other statistics used in the meta-analysis (thus violating the usual assumption of nonindependence in meta-analytic research). Of the remaining 15 statistics, half were based on studies that have been previously evaluated by Shields (1978) as having serious methodological flaws, with only 6 using outcome ratings by raters unaware of the Rorschach data. Of these 6 studies, the sole study using an independent rating (rather than the therapist's rating) for treatment outcome had a very small sample ( $N = 11$ ; Filmer-Bennett, 1955). We echo Meyer and Handler's call for contemporary researchers to investigate the merits of the RPRS, as little research on this scale has been conducted since the early 1980s. However, it will be important for such research to address both Shield's (1978) concern about the scale's sensitivity and the potential confound stemming from the strong association between the RPRS and intelligence ( $r = .66$ ; Hathaway, 1982).

Even accepting a modest relation between certain Rorschach scales and treatment outcome, there remains the question of how this information benefits psychologists and, ultimately, clients. Why is it useful to know that certain clients are expected to improve more than others in therapy? Are the modest associations that may exist sufficient in magnitude to justify refusing treatment to those predicted to be more likely to leave treatment or to improve less? Without explicit consideration of questions like these, it is not possible to evaluate the Rorschach's clinical utility as a predictor of treatment response variables.

Our analysis of the literature dealing with the relevance of the Rorschach to therapy services is that, even if the Rorschach could provide valid information about personality dynamics, personality structure, and treatment outcome, there is currently no replicated evidence to indicate that this information has any meaningful bearing on services provided to clients and no evidence to substantiate claims about improved treatment outcome accruing from this information. Such a conclusion, when combined with (a) limited reliability and validity data and (b) the costly nature of Rorschach assessments (i.e., approximately 3 hr of clinician time), has direct implications for clinical practice. There is currently insufficient evidence regarding the validity and clinical utility of each Comprehensive System score to support the widespread use of Rorschach scales in legal, forensic, or occupational assessments. At present, as there is insufficient evidence indicating that Comprehensive System or other Rorschach scales can reliably predict treatment outcome, there is no scientific reason to justify the use of Rorschach data in psychotherapy planning.

### Conclusions

The frequent explanations offered over the years to account for the limited research support of the Rorschach, including discus-

sions of power, research design, and problems in the selection of appropriate validation criteria, are not substitutes for empirical evidence. Implicit in such discussions is the assumption that the Rorschach has clinical utility and that researchers have merely not yet been able to demonstrate this obvious fact. After several decades of post hoc explanations for the lack of empirical support, the time has come to accept that neither the Rorschach (as a generic clinical test or method) nor the Comprehensive System has a firm enough basis to warrant widespread clinical use. Although some Rorschach scales appear to have validity in basic research, their practical clinical utility has not yet been demonstrated.

At this juncture, it is instructive to consider the conclusions reached by some Rorschach supporters 20 years ago. In reviewing the scientific status of the Rorschach, Schwartz and Lazar (1979) identified the key issue dividing opponents and proponents of the Rorschach as a fundamental difference in their methods of describing human behavior. At odds, they argued, were two different worldviews, one espousing the use of statistics and norms, the other advocating a focus on meaning and clinical experience. For their part, Aronow, Reznikoff, and Rauchway (1979) argued that the dual nature of the Rorschach, as a test amenable to both nomothetic and ideographic analysis, must be accepted if progress was to occur. Without such acceptance, they sagely predicted that the next 20 years would continue to be characterized by disagreements between Rorschach advocates and critics.

The Comprehensive System is an attempt to bridge these the two worldviews described by these commentators. Twenty years later, it appears that the differing world views may ultimately be incompatible. Exner's efforts to systematize the Rorschach and to meet professional standards for psychological tests are laudable, but the Comprehensive System does not yet meet these standards and cannot meet these standards until there exists replicated, peer-reviewed evidence supporting the reliability, validity, and utility of every scale included in the Comprehensive System. Because of the enormity of the undertaking this would require, we doubt that it will ever occur. However, for those interested in undertaking such an enterprise, numerous sources are available to assist in designing and analyzing Rorschach research (e.g., Exner, 1995; Meyer, 1996a; Shontz & Green, 1992; Viglione, 1997; Weiner, 1995; Widiger & Schilling, 1980).

Available evidence does support the continued use in research contexts of some specific scales, including some that are not included in the Comprehensive System. Given the meager support from thousands of publications to date, the history of disagreements among proponents about the proper use of the Rorschach, and the uncertain acceptance by psychologists of a psychometrically and scientifically sound approach to Rorschach scoring and interpretation, we doubt that there will ever be sufficient evidence to suggest that the Rorschach or the Comprehensive System can contribute, in routine clinical practice, to scientifically informed psychological assessment.

### References

- Ackerman, M. J. (1995). *Clinician's guide to child custody evaluations*. New York: Wiley.
- Ackerman, M. J., & Ackerman, M. C. (1997). Custody evaluation practices: A survey of experienced professionals (revisited). *Professional Psychology: Research and Practice*, 28, 137-145.

- Acklin, M. W. (1993). Integrating the Rorschach and the MMPI in clinical assessment: Conceptual and methodological issues. *Journal of Personality Assessment, 60*, 125–131.
- Acklin, M. W., McDowell, C. J., & Orndoff, S. (1992). Statistical power and the Rorschach: 1975–1991. *Journal of Personality Assessment, 59*, 366–379.
- Allard, G., Butler, J., Faust, D., & Shea, M. T. (1995). Errors in hand scoring objective personality tests: The case of the Personality Diagnostic Questionnaire—Revised (PDQ-R). *Professional Psychology: Research and Practice, 26*, 304–308.
- Anderson, T., & Dixon, W. E. (1993). The factor structure of the Rorschach for adolescent inpatients. *Journal of Personality Assessment, 60*, 319–333.
- Archer, R. P. (1996). MMPI-Rorschach interrelationships: Proposed criteria for evaluating explanatory models. *Journal of Personality Assessment, 67*, 504–515.
- Archer, R. P., & Gordon, R. A. (1988). MMPI and Rorschach indices of schizophrenic and depressive diagnoses among adolescent inpatients. *Journal of Personality Assessment, 52*, 276–287.
- Archer, R. P., & Krishnamurthy, R. (1993a). A review of MMPI and Rorschach interrelationships in adult samples. *Journal of Personality Assessment, 61*, 277–293.
- Archer, R. P., & Krishnamurthy, R. (1993b). Combining the Rorschach and the MMPI in the assessment of adolescents. *Journal of Personality Assessment, 60*, 132–140.
- Archer, R. P., & Krishnamurthy, R. (1997). MMPI-A and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment, 69*, 517–533.
- Aronow, E., & Reznikoff, M. (1972). Attitudes toward the Rorschach test expressed in book reviews: A historical perspective. *Journal of Personality Assessment, 37*, 309–315.
- Aronow, E., Reznikoff, M., & Moreland, K. L. (1995). The Rorschach: Projective technique or psychometric test? *Journal of Personality Assessment, 64*, 213–228.
- Aronow, E., Reznikoff, M., & Rauchway, A. (1979). Some old and new directions in Rorschach testing. *Journal of Personality Assessment, 43*, 227–234.
- Atkinson, L. (1986). The comparative validities of the Rorschach and MMPI: A meta-analysis. *Canadian Psychology, 27*, 238–247.
- Atkinson, L., Quarrington, B., Alp, I. E., & Cyr, J. J. (1986). Rorschach validity: An empirical approach to the literature. *Journal of Clinical Psychology, 42*, 360–362.
- Ball, J. D., Archer, R. P., Gordon, R. A., & French, J. (1991). Rorschach depression indices with children and adolescents: Concurrent validity findings. *Journal of Personality Assessment, 57*, 465–476.
- Ball, J. D., Archer, R. P., & Imhof, E. A. (1994). Time requirements of psychological testing: A survey of practitioners. *Journal of Personality Assessment, 63*, 239–249.
- Blais, M. A., Norman, D. K., Quintar, B., & Herzog, D. B. (1995). The effect of administration method: A comparison of the Rapaport and Exner Rorschach systems. *Journal of Clinical Psychology, 51*, 100–107.
- Blatt, S. J. (1975). The validity of projective techniques and their research and clinical contribution. *Journal of Personality Assessment, 39*, 327–343.
- Bornstein, R. F. (1996). Construct validity of the Rorschach Oral Dependency scale: 1967–1995. *Psychological Assessment, 8*, 200–205.
- Butcher, J., & Rouse, S. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology, 47*, 87–111.
- Carlson, C. F., Kula, M. L., & St. Laurent, C. M. (1997). Rorschach revised DEPI and CDI with inpatient major depressives and borderline personality disorder with major depression: Validity issues. *Journal of Clinical Psychology, 53*, 51–58.
- Cronbach, L. J. (1949). Statistical methods applied to Rorschach scores: A review. *Psychological Bulletin, 46*, 393–429.
- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: The Free Press.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668–1674.
- DeCato, C. M. (1993). On the Rorschach M response and monotheism. *Journal of Personality Assessment, 60*, 362–378.
- Durand, V., Blanchard, E., & Mindell, J. (1988). Training in projective testing: Survey of clinical training directors and internship directors. *Professional Psychology: Research and Practice, 19*, 236–238.
- Elliott, A. J., Miltenberger, R. G., Kaster-Bundgaard, J., & Lumley, V. (1996). A national survey of assessment and therapy techniques used by behavior therapists. *Cognitive and Behavioral Practice, 3*, 107–125.
- Exner, J. E. (1974). *The Rorschach: A comprehensive system: Vol. 1*. New York: Wiley.
- Exner, J. E. (1986). *The Rorschach: A comprehensive system: Vol. 1 Basic foundations* (2nd ed.). New York: Wiley.
- Exner, J. E. (1988). Problems with brief Rorschach protocols. *Journal of Personality Assessment, 52*, 640–647.
- Exner, J. E. (1989). Searching for projection in the Rorschach. *Journal of Personality Assessment, 53*, 520–536.
- Exner, J. E. (1991). *The Rorschach: A comprehensive system: Vol. 2. Interpretation* (2nd ed.). New York: Wiley.
- Exner, J. E. (1992). R in Rorschach research: A ghost revisited. *Journal of Personality Assessment, 58*, 245–251.
- Exner, J. E. (1993). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations* (3rd ed.). New York: Wiley.
- Exner, J. E. (Ed.). (1995). *Issues and methods in Rorschach research*. Hillsdale, NJ: Erlbaum.
- Exner, J. E. (1997). The future of Rorschach in personality assessment. *Journal of Personality Assessment, 68*, 37–46.
- Exner, J. E., & Exner, D. E. (1972). How clinicians use the Rorschach. *Journal of Personality Assessment, 36*, 403–408.
- Exner, J. E., Viglione, D. J., & Gillespie, R. (1984). Relationships between Rorschach variables as relevant to the interpretation of structural data. *Journal of Personality Assessment, 48*, 65–70.
- Filmer-Bennett, G. (1955). The Rorschach as a means of predicting treatment outcome. *Journal of Consulting and Clinical Psychology, 19*, 331–334.
- Fischer, C. T. (1994). Rorschach scoring questions as access to dynamics. *Journal of Personality Assessment, 62*, 515–524.
- Ganellen, R. J. (1996a). Comparing the diagnostic efficiency of the MMPI, MCMI-II, and the Rorschach: A review. *Journal of Personality Assessment, 67*, 219–243.
- Ganellen, R. J. (1996b). Exploring MMPI-Rorschach relationships. *Journal of Personality Assessment, 67*, 529–542.
- Ganellen, R. J. (1996c). *Integrating the Rorschach and the MMPI-2 in personality assessment*. Hillsdale, NJ: Erlbaum.
- Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review, 4*, 641–655.
- Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387–396.
- Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.
- Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the MMPI: Results from meta-analyses. *Psychological Science, 9*, 402–404.
- Garfield, S. L., & Kurtz, R. M. (1973). Attitudes toward training in diagnostic testing: A survey of directors of internship training. *Journal of Consulting and Clinical Psychology, 40*, 350–355.
- Greenwald, D. F. (1990). An external construct validity study of Rorschach personality variables. *Journal of Personality Assessment, 55*, 768–780.

- Groth-Marnat, G. (1997). *Handbook of psychological assessment* (3rd ed.). New York: Wiley.
- Hathaway, A. P. (1982). Intelligence and non-intelligence factors contributing to scores on the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment*, 46, 8–11.
- Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist*, 42, 963–974.
- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hilsenroth, M. J., & Handler, L. (1995). A survey of graduate students' experiences, interests, and attitudes about learning the Rorschach. *Journal of Personality Assessment*, 64, 243–257.
- Hilsenroth, M. J., Handler, L., Toman, K. M., & Padawer, J. R. (1995). Rorschach and MMPI-2 indices of early psychotherapy termination. *Journal of Consulting and Clinical Psychology*, 63, 956–965.
- Howes, R. J. (1981). The Rorschach: Does it have a future? *Journal of Personality Assessment*, 45, 339–351.
- Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, 2, 447–472.
- Jackson, C. W., & Wohl, J. (1966). A survey of Rorschach teaching in the university. *Journal of Projective Techniques and Personality Assessment*, 30, 115–134.
- Jensen, A. R. (1965). A review of the Rorschach. In O. K. Buros (Ed.), *Sixth mental measurement yearbook* (pp. 501–509). Highland Park, NH: Gryphon.
- Kendell, R. E. (1989). Clinical validity. *Psychological Medicine*, 19, 45–55.
- Khadivi, A., Wetzler, S., & Wilson, A. (1997). Manic indices on the Rorschach. *Journal of Personality Assessment*, 69, 365–375.
- Kinder, B. N. (1992). The problems of R in clinical settings and in research: Suggestions for the future. *Journal of Personality Assessment*, 58, 252–259.
- Kinder, B., Brubaker, R., Ingram, R., & Reading, E. (1982). Rorschach form quality: A comparison of the Exner and Beck systems. *Journal of Personality Assessment*, 46, 131–138.
- Kleiger, J. H. (1992). A conceptual critique of the EA:es comparison in the Comprehensive Rorschach System. *Psychological Assessment*, 4, 288–296.
- Klopfer, B., Kirkner, F., Wisham, W., & Baker, G. (1951). Rorschach Prognostic Rating Scale. *Journal of Projective Techniques and Personality Assessment*, 15, 425–428.
- Kramer, R. L. (1991). The Rorschach M response: A return to its roots. *Journal of Personality Assessment*, 57, 30–36.
- Lees-Haley, P. R. (1992). Psychodiagnostic test usage by forensic psychologists. *American Journal of Forensic Psychology*, 10, 25–30.
- Levy, L. H., & Orr, T. B. (1959). The social psychology of Rorschach validity research. *Journal of Abnormal and Social Psychology*, 58, 79–83.
- Lipgar, R. M. (1992). The problem of R in the Rorschach: The value of varying responses. *Journal of Personality Assessment*, 58, 223–230.
- Lipovsky, J. A., Finch, A. J., & Belter, R. W. (1989). Assessment of depression in adolescents: Objective and projective measures. *Journal of Personality Assessment*, 53, 449–458.
- Lubin, B., Larsen, R., & Matarazzo, J. (1984). Patterns of psychological test usage in the United States. *American Psychologist*, 39, 451–453.
- Lubin, B., Wallis, R., & Paine, C. (1971). Patterns of psychological test usage in the United States: 1935–1969. *Professional Psychology*, 2, 70–74.
- Masling, J. M. (1960). Interpersonal and situational influence in projective testing. *Psychological Bulletin*, 57, 65–85.
- Masling, J. M. (1992). The influence of situation and interpersonal variables in projective testing. *Journal of Personality Assessment*, 59, 616–640.
- Mason, B. J., Cohen, J. B., & Exner, J. E. (1985). Schizophrenic, depressive, and nonpatient personality organizations described by Rorschach factor structures. *Journal of Personality Assessment*, 49, 295–305.
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Bulletin*, 96, 690–702.
- McCully, R. S. (1965). Current attitudes about projective techniques in APA approved internship centers. *Journal of Projective Techniques and Personality Assessment*, 27, 271–280.
- McDowell, C., & Acklin, M. W. (1996). Standardizing procedures for calculating Rorschach interrater reliability: Conceptual and empirical foundations. *Journal of Personality Assessment*, 66, 308–320.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52, 194–216.
- Meyer, G. J. (1992a). Response frequency problems in the Rorschach: Clinical and research implications with suggestions for the future. *Journal of Personality Assessment*, 58, 231–244.
- Meyer, G. J. (1992b). The Rorschach's factor structure: A contemporary investigation and historical review. *Journal of Personality Assessment*, 59, 117–136.
- Meyer, G. J. (1993). The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI-2 criteria. *Journal of Personality Assessment*, 60, 153–180.
- Meyer, G. J. (1996a). The Rorschach and MMPI: Toward a more scientifically differentiated understanding of cross-method assessment. *Journal of Personality Assessment*, 67, 558–578.
- Meyer, G. J. (1996b). Construct validation of scales derived from the Rorschach method: A review of issues and introduction to the Rorschach Rating Scale. *Journal of Personality Assessment*, 67, 598–628.
- Meyer, G. J. (1997a). Assessing reliability: Critical corrections for a critical examination of the Rorschach Comprehensive System. *Psychological Assessment*, 9, 480–489.
- Meyer, G. J. (1997b). Thinking clearly about reliability: More critical corrections regarding the Rorschach Comprehensive System. *Psychological Assessment*, 9, 495–498.
- Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment*, 69, 1–38.
- Nezworski, M. T., & Wood, J. M. (1995). Narcissism in the Comprehensive System for the Rorschach. *Clinical Psychology: Science and Practice*, 2, 179–199.
- Nichols, D. S. (1996). Remarks on method for MMPI-Rorschach studies. *Journal of Personality Assessment*, 67, 516–528.
- Norcross, J. C., Karg, R. S., & Prochaska, J. O. (1997). Clinical psychologists in the 1990s: Part II. *The Clinical Psychologist*, 50, 4–11.
- Omberg, B., & Zalewski, C. (1994). Assessment of adolescents with the Rorschach: A critical review. *Assessment*, 1, 209–217.
- Parker, K. (1983). A meta-analysis of the reliability and validity of the Rorschach. *Journal of Personality Assessment*, 47, 227–231.
- Parker, K. C. H., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin*, 103, 367–373.
- Piotrowski, C., & Keller, J. W. (1984). Attitudes toward clinical assessment by members of the AABT. *Psychological Reports*, 55, 831–838.
- Piotrowski, C., Sherry, D., & Keller, J. W. (1985). Psychodiagnostic test usage: A survey of the Society for Personality Assessment. *Journal of Personality Assessment*, 49, 115–119.
- Piotrowski, C., & Zalewski, C. (1993). Training in psychodiagnostic testing in APA-approved PsyD and PhD clinical psychology programs. *Journal of Personality Assessment*, 61, 394–405.
- Pruitt, J. A., Smith, M. C., Thelen, M. H., & Lubin, B. (1985). Attitudes of

- academic clinical psychologists toward projective techniques: 1968–1983. *Professional Psychology: Research and Practice*, 16, 781–788.
- Ritzler, B., & Alter, B. (1986). Rorschach teaching in APA-approved clinical graduate programs: Ten years later. *Journal of Personality Assessment*, 50, 44–49.
- Ritzler, B. A., & Del Gaudio, A. C. (1976). A survey of Rorschach teaching in APA-approved clinical graduate programs. *Journal of Personality Assessment*, 40, 451–453.
- Ritzler, B., & Nalesnik, D. (1990). The effect of inquiry on the Exner Comprehensive System. *Journal of Personality Assessment*, 55, 647–656.
- Schwartz, F., & Lazar, J. (1979). The scientific status of the Rorschach. *Journal of Personality Assessment*, 43, 3–11.
- Shields, R. B. (1978). The usefulness of the Rorschach Prognostic Rating Scale—A rebuttal. *Journal of Personality Assessment*, 42, 579–582.
- Shontz, F. C., & Green, P. (1992). Trends in research on the Rorschach: Review and recommendations. *Applied & Preventive Psychology*, 1, 149–156.
- Smith, B., & Sechrest, L. (1991). Treatment of aptitude  $\times$  treatment interactions. *Journal of Consulting and Clinical Psychology*, 59, 233–244.
- Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, 112, 140–154.
- Standards for educational and psychological testing*. (1985). Washington, DC: American Psychological Association.
- Thelen, M. H., Varble, D. L., & Johnson, J. (1968). Attitudes of academic clinical psychologists toward projective techniques. *American Psychologist*, 23, 517–521.
- Turk, D. C., & Salovey, P. (Eds.). (1988). *Reasoning, inference, and judgment in clinical psychology*. New York: The Free Press.
- Viglione, D. J. (1996). Data and issues to consider in reconciling self-report and the Rorschach. *Journal of Personality Assessment*, 67, 579–587.
- Viglione, D. J. (1997). Problems in Rorschach research and what to do about them. *Journal of Personality Assessment*, 68, 590–599.
- Vincent, K. R., & Harman, M. J. (1991). The Exner Rorschach: An analysis of clinical validity. *Journal of Clinical Psychology*, 47, 596–599.
- Wade, T. C., & Baker, T. B. (1977). Opinions and use of psychological tests: A survey of clinical psychologists. *American Psychologist*, 32, 874–882.
- Wade, T. C., Baker, T. B., Morton, T. L., & Baker, L. J. (1978). The status of psychological testing in clinical psychology: Relationships between test use and professional activities and orientations. *Journal of Personality Assessment*, 42, 3–10.
- Watkins, C. E., Jr., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice*, 26, 54–60.
- Weiner, I. B. (1986). Conceptual and empirical perspectives on the Rorschach assessment of psychopathology. *Journal of Personality Assessment*, 50, 472–479.
- Weiner, I. B. (1991). Editor's note: Interscorer agreement in Rorschach research. *Journal of Personality Assessment*, 56, 1.
- Weiner, I. B. (1993). Clinical considerations in the conjoint use of the Rorschach and the MMPI. *Journal of Personality Assessment*, 60, 148–152.
- Weiner, I. B. (1994). The Rorschach Inkblot Method (RIM) is not a test: Implications for theory and practice. *Journal of Personality Assessment*, 62, 498–504.
- Weiner, I. B. (1995). Methodological considerations in Rorschach research. *Psychological Assessment*, 7, 330–337.
- Weiner, I. B. (1996). Some observations on the validity of the Rorschach Inkblot Method. *Psychological Assessment*, 8, 206–213.
- Weiner, I. B. (1997). Current status of the Rorschach Inkblot Method. *Journal of Personality Assessment*, 68, 5–19.
- Widiger, T. A., & Schilling, K. M. (1980). Toward a construct validation of the Rorschach. *Journal of Personality Assessment*, 44, 450–459.
- Wood, J. M., Nezworski, T., & Stejskal, W. J. (1996a). The comprehensive system for the Rorschach: A critical examination. *Psychological Science*, 7, 3–10.
- Wood, J. M., Nezworski, T., & Stejskal, W. J. (1996b). Thinking critically about the comprehensive system for the Rorschach: A reply to Exner. *Psychological Science*, 7, 14–17.
- Wood, J. M., Nezworski, T., & Stejskal, W. J. (1997). The reliability of the Comprehensive System for the Rorschach: A comment on Meyer (1997). *Psychological Assessment*, 9, 490–494.

Received March 4, 1998  
 Revision received June 7, 1998  
 Accepted April 19, 1999 ■